The Dissertation Committee for Xiaoshu Han
certifies that this is the approved version of the following dissertation:

Dynamics of Health and Employment: Theory, Evidence and Policy Implications

Committee:

Russell W. Cooper, Supervisor

Jerome Adda

Dean Corbae

Stephen Donald

Burhanettin Kuruscu

Stephen Trejo

Dynamics of Health and Employment: Theory, Evidence and Policy Implications

by

Xiaoshu Han, B.A.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2007

Dedicated to my mother Mingshen, my husband Harry and my daughter Erica.

Acknowledgments

I wish to thank the multitudes of people who helped me. I am grateful to my thesis supervisor Russell Cooper for his support and guidance. I would like to thank Stephen Bronars, Dean Corbae, Stephen Donald, Burhanettin Kuruscu, Gerald Oettinger, John Rust and Stephen Trejo for their valuable comments and suggestions. I also thank seminar participants at the Society for Economic Dynamics 2006 conference in Vancouver, Econometrics Society North American 2006 Summer Meeting in Minneapolis and the Southern Economic Association 2006 conference in Charleston for helpful comments. I appreciated the helpful comments from all the participants in the labor lunch seminar and macroeconomics writing course at University of Texas at Austin. All the errors remain with author.

Dynamics of Health and Employment: Theory, Evidence and Policy Implications

Publication No. __________________

Xiaoshu Han, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Russell W. Cooper

This study involves the relationships between health, medical expenditure and labor force participation, and implement these relationships in policy evaluations.

The first chapter examines the relationships between health, labor force participation and medical spending in a simultaneous equations model using data from the Medical Expenditure Panel Survey (MEPS). It finds a significant positive effect of health on labor force participation. The estimation results of medical spending equation show that healthier individuals spend less on their medical care and surprisingly, employment status has no significant effect on medical spending. Interestingly, employment status does impact health status positively. The most puzzling result is the significant negative coefficient on medical spending in the health equation.

The puzzling result motivates this dynamic heterogeneous agent model of Chapter Two to study the relationships between health status, medical expenditure and employment. In this model, individuals value both consumption and health and they are heterogeneous in their health levels and employment status. Individuals choose how much to spend on medical care and consumption. Health can be accumulated from investment in medical care and this increases both job opportunities and quality of life. The structural parameters are estimated by an indirect inference procedure which matches the simulated regression coefficients to the data regression coefficients from the Medical Expenditure Panel Survey. This chapter finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure impacts the health in the positive way. This paper also explains the basis for this result.

Chapter Three is a policy application of the model developed in chapter Two. It builds on the model of Chapter Two and adds the employer-provided health insurance. It concentrates on measuring the welfare cost of mandating employer provided health insurance. It compares the welfare of the working-aged individuals before and after the mandate and finds a welfare loss of 0.7 percent of GDP.

Table of Contents

List of Tables

List of Figures

Chapter 1

Health Status, Medical Spending and Labor Force Participation

This chapter examines the relationships between health, labor force participation and medical spending in a simultaneous equations model using data from the Medical Expenditure Panel Survey (MEPS). It finds a significant positive effect of health on labor force participation. The estimation results of medical spending equation show that healthier individuals spend less on their medical care and surprisingly, employment status has no significant effect on medical spending. Interestingly, employment status does impact health status positively. The most puzzling result is the significant negative coefficient on medical spending in the health equation.

1.1 Introduction

The relationship between health status and labor force participation has been explored extensively in the literature. The relationship is important for policy makers to know in order to assess the effectiveness and solvency of the employment and health related policies. Most of the previous researches simply considers how health impacts labor force participation and suggests poor health reduces the capacity to work and has substantive effects on labor force participation. Currie and Madrian (1999) has a detailed review of them. But since health status and labor force participation have mutual effects of each other, a simultaneous equations model has been used by Stern (1989), followed by Cai and Kalb (2004) to endogenize both factors.

For the previous literature, almost all of them focus on the effects of elderly working individuals' health on their labor force participation decisions which is motivated by the increase of early retirement of older men. Cai and Kalb (2004) is an exception and explores the effect for individuals at all working ages in an Australian context. This paper focuses on the individuals at all working ages but in the U.S. context. Furthermore, no one in the literature of the relationship between health status and labor force participation addresses how medical care expenditures impact health.

As we all know, medical care expenditure in U.S. now is almost 15 percent of its total GDP (Source: CMS, Office of the Actuary, National Health Statistics Group.). It must have played a big role in people's health. Rosenzweig and Schultz (1983, 1988, 1991), Corman et al.(1987), Grossman and Joyce(1990), and Joyce(1994) consider the effects of some health inputs such as mother's prenatal care and smoking behavior on baby's health. They use baby's weight as a measure of baby's health and use income, education and price of medical care service as instruments for medical care. Stratmann (1999) estimates the effect of doctor visits on work day loss using the types of health insurance as instruments for doctor visits and finds negative coefficients on doctor visits. In that study, the work day loss is the loss due a certain health condition. Three different types of health conditions used in three separate regressions are influenza, impairments and asthma. And doctors' visits are those visits which contribute to the treatment of those conditions. No sufficient work has been done to evaluate directly how medical care expenditures impact health for the working age individuals. This paper considers the effect of out-of-pocket medical expenses of working age people on their health.

The relationships among health status, medical expenses and labor force participation are examined in a simultaneous equations model in this paper using Medical Expenditure Panel Survey (MEPS). The paper is arranged as follows: Section 2 discusses different health measures used in the literature and the health measure used in this paper. Section 3 outlines the modeling strategy. Section 4 describes the data and model specifications. Section 5 presents the estimation results. Section 6 concludes the paper and discusses a certain unsolved issues in this paper which deserves future work.

1.2 Measure of Health

In the literature of the relationship between health status and labor force participation, there are extensive debates on what kinds of measure to use for health: whether it should be an objective measure such as the presence of chronic and acute conditions and nutritional status(e.g., height, weight, or body mass index (BMI)) or it should be a subjective measure such as self-assessed health status derived from individuals' responses to survey questions:

"would you describe your health as excellent, very good, good, fair, or poor?" or "Does your health limit the amount or kind of work you can do". For the objective measure, indicators for specific conditions may not be very directly related to ones' productivity (Currie and Madrian, 1999). Individuals who lost one limb may be highly productive as a typist. At the mean time, the objective measure suffers the measurement error too. As for the subjective measures, they may be more directly related to productivity but they may also be more subject to reporting biases. Individuals who have exited the labor force may be more likely to report that they have poor health status and functional limitations to rationalize their behavior. Currie and Madrian (1999) and Cai and Kalb (2004) both have detailed discussions of different measures. When self-reported measures are used, health seems to play a larger role and economic factors a smaller one on the labor force participation than when more objective measures are used (Bound 1991). While self-reported health measure still remains a popular measure for health, it is still an open question as for which measure is the best to use.

The health measure used in this paper is the Physical Component Summary (PCS) scores formed from the answers to the Short-Form 12 questions (Table 1.1). There are both physical and mental health questions in these Short-Form 12 questions. There are two forms of summary scores: PCS and MCS (Mental Component Summary).[1] PCS puts more weight on physical

[1] As for how to calculate these summary scores, please refer to 'Ware,Jr., J.E., Kosinski, M., and Keller, S. How to Score the SF-12(r)Physical and Mental Health Summary Scales

health questions than MCS does. The PCS scores have several advantages over the self-reported health measure mentioned above.

First, the reason why researchers concern about the rationalization problem in the self-reported health measure is that most of the existing literature is on the elderly working age individuals. Some of them retire early because of their poor health but the others exit the labor force early because of other reasons such as leisure since they already have enough savings. If they retire earlier than 65 then they are not eligible for medicare and have to pay for the expensive health insurance. It is highly possible for them to go to doctor's office more often and report poor health in order to qualify for the disability health insurance which is covered by the social security insurance. However, this paper concentrates on all working age individuals and the elderly is only a small proportion of it so it is probably not such a big issue.

Second, the self-assessed health status used in the existing literature all comes from the same set of interview as the other information such as employment, income and demographic characteristics. Individuals assess their health after they are asked about the employment and other related questions. It is highly possible for them to rationalize their answers to health questions. The Short-Form 12 questions used in this paper is from a self-administered questionnaire (SAQ), a paper-and-pencil questionnaire, delivered to the interviewers by mail. The SAQ is designed to collect a variety of health status and

(Third Edition).(September 1998) QualityMetric,Inc., Lincoln,RI.'. This manual can be purchased from QualityMetric,Inc.

health care quality measures from adults and it is separated and conducted at different time from the personal interview which has the employment, income and other information. Individuals may be more candid to sensitive questions, such as health status, when filling out a self-completion form rather than being interviewed (Tourangeau and Smith, 1996) and knowing that it is only a health care quality survey.

Third, the self-assessed health status used in the existing literature is too general to compare across different individuals. Individuals always feel good about themselves tend to give more optimistic answers. Most of the Short-Form 12 questions are very specific questions which are very straight forward to answer. In addition, there is a mental health component in the PCS scores.

Lastly, the PCS and MCS scores have been tested by health professionals. Hurst et al. (1998) test them on rheumatoid arthritis (RA) patients and find that the SF12 is a reliable, valid and responsive measure of health status in the majority of RA patients, and meets standards required for comparing groups of patients. Melville et al. (2003) find that quality of life assessment using the short form 12 questionnaire is as reliable and sensitive as the short form 36 in distinguishing symptom severity in myocardial infarction survivors. Luo et al. (2003) find that the short form 12-item survey demonstrated good internal consistency reliability, construct validity, and responsiveness in patients with back pain. The most thorough and extensive research on the validity of the PCS and MCS scores has been conducted by Avery et al. (2004). The research tests the scores on various chronic conditions such as diabetes, arthritis, heart disease, stroke, cancer and so on and different risk factors such as high blood pressure, smoking status, body mass index and so on against different population groups.

1.3 The Model and Methods of Estimation

Some of the researchers explore the mutual effects of health and labor force participation but none of them take into consideration of the impact of individuals' medical spending on their health. But if the medical spending is considered in the health equation, it becomes endogenous too since individuals' health and employment status also have impacts on their medical spending. The model in this paper is a simultaneous equations model of health status, labor force participation and medical expenditure. It takes the following forms:

$$\begin{aligned} h &= \alpha_1 l + \alpha_2 m + X_0\alpha_3 + X_1\alpha_4 + \varepsilon_1 \\ l &= \beta_1 h + X_0\beta_2 + X_2\beta_3 + \varepsilon_2 \\ m &= \gamma_1 h + \gamma_2 l + X_0\gamma_3 + X_3\gamma_4 + \varepsilon_3 \end{aligned}$$

where h, l and m are endogenous variables, h is individual's health, l is individual's labor force participation decision and m is individual's out-of-pocket medical expenditure. X_0, X_1, X_2 and X_3 are exogenous variables; X_1, X_2 and X_3 are excluded variables from the other equations. By assumption, $E(\varepsilon_1 Z) = 0,\ E(\varepsilon_2 Z) = 0,\ E(\varepsilon_3 Z) = 0$, and $Z = (X_0\ X_1\ X_2\ X_3)$.

Table 1.1: The SF-12 Questions:

General health today	poor, fair, good, very good, excellent
During a typical day, limitations in moderate activities	limited a lot, limited a little, not limited
During a typical day, limitations in climbing several flights of stairs	limited a lot, limited a little, not limited
During past 4 weeks, as result of physical health, accomplished less that would like	yes/no
During past 4 weeks, as result of physical health, limited in kind of work or other activities	yes/no
During past 4 weeks, as result of mental problems, accomplished less that would like	yes/no
During past 4 weeks, as result of mental problems, limited in kind of work or other activities	yes/no
During past 4 weeks, pain interfered with normal work outside the home and housework	not at all, a little bit, moderately, quite a bit, extremely
During the past 4 weeks, felt calm and peaceful	all the time, most of the time, good bit of time,some of the time, little of the time, none of the time
During the past 4 weeks, had a lot of energy	Same As Above
During the past 4 weeks, felt downhearted and blue	Same As Above
During past 4 weeks, physical health or emotional problems interfered with social activities	Same As Above

The model can be consistently estimated by a two step procedure proposed by Maddala (1983). The two step procedure is similar to two stage least squares but with the possibility of allowing for the nonlinear regressions. In the first step, each endogenous variable in the system is regressed on all the exogenous variables. Then in the second step, we estimate each equation in the system replacing the endogenous variables with their fitted value , and from the first step.

1.4 Data and model specification

1.4.1 The data

The data used for this paper come from the Household Component of the Medical Expenditure Panel Survey. The MEPS HC is a nationally representative survey of the U.S. civilian noninstitutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a -year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures. MEPS HC panel 6 covers two years' data (2001 and 2002) for 21,959 individuals. Since this paper's main goal is to explore the relationships for the working age individuals, the sample of individuals whose ages are either below 18 or over 65 is dropped. Students are out of sample too. After cleaning the sample with the missing information, 8896 data points are left.

1.4.2 Model specification

In this subsection all the variables in the three simultaneous equations are described. Table 1.2 provides the definitions.

In order to identify the simultaneous equation models, there have to be at least as many excluded exogenous variables as endogenous variables in the equation theoretically (Maddala, 1983). For example, in the health equation, there are two endogenous variables on the right hand side: employment status and medical spending since both of them have impacts on individuals' health. In order to identify the health equation, we have to find at least two exogenous variables which don't appear in the health equation but appear in either labor force participation equation or medical expenditure equation.

Three endogenous variables are health, employment status and medical spending. The health measure is the PCS score as mentioned in Section 1.2. The score ranges from 0 to 100 where 100 corresponds to the highest level of health. The lowest health level in the current sample is 11.73, the highest is

67.24 and the mean is 50.07 (Refer Table 1.3 for descriptive statistics). The employment status takes two values: if the individual is employed, then it is 1, otherwise it is 0. Medical spending in this paper is individual's out-of-pocket spending. It ranges from 0 to 37128 dollars per year and the average is 507 dollars per year.

The common exogenous variables in all three equations are individual's age, sex, race, marital status and education level. The appearance of there variables are standard in the literature.

Additional exogenous variables in the health equation are smoke, physical activity, seat belt and individuals' past health. Individuals' smoking behavior and whether they exercise regularly have long been considered to have direct impacts on health in the literature. Whether individuals wear seat belt most of the times only affect the health when there are car accidents. Individuals' health depend greatly on their preexisting conditions which are their last period's health level. They are assumed to have impacts on labor force participation and medical spending only through their impacts on the current health level.

Additional exogenous variables in the labor force participation equation are age square, number of children, whether the individual has children under 5 years old, whether the individual is employed in the last period and spouse' income. The variables used here are quite standard in the literature of labor supply.

Additional exogenous variables in the medical spending equation are different health insurance types and spouse' income. The different types of insurance and spouse's income only impact the health through their impact on medical spending. I assume here the different types of health insurance are uncorrelated with the employment status because unemployed individuals have access to public insurance such as Medicaid and nowadays, Medicaid has both HMO and fee for service.

Table 1.2: Variable definitions:

Endogenous variables	
$health_t$	physical component summary index at period t, 0 is the lowest heath level, 100 is the highest
$employed_t$	employment status at period t:1 if in labor force, 0 otherwise
$medical_t$	total amount of out-of-pocket medical expenditure
Variables appearing in all the equations	
age	Individual's actual age
sex	1 if male, 0 otherwise
race	1 if white, 0 otherwise
mar	1 if married, 0 otherwise
edu	1 if with a college degree, 0 otherwise
Additional variables appearing in the health equation	
smoke	1 if smoke, 0 otherwise
phyact	1 if currently spends half hour or more on moderate to vigorous physical activities at least three times a week, 0 otherwise
seatbelt	wears seat belt whenever drives or rides in a car: 1 if always, 0 otherwise
$health_{t-1}$	physical component summary index at period t-1, 0 is the lowest heath level, 100 is the highest
Additional variables appearing in the labor force participation equation	
$age2$	age^2
numchi	the actual number of children the individual has
$child5$	whether the individual has children under 5 years old: 1 if has, 0 otherwise
$employed_{t-1}$	employment status at period t-1
spousein	spouse' actual income: if no spouse then spousein=0
Additional variables appearing in the medical expenditure equation	
hmo	the individual's health insurance type: 1 if hmo, 0 otherwise
gatekeeper	the individual's health insurance type: 1 if gatekeeper, 0 otherwise
ppo	the individual's health insurance type: 1 if ppo, 0 otherwise
spousein	spouse' actual income: if no spouse then spousein=0

Table 1.3: Descriptive statistics of the sample:

Variables	Sample Mean	Standard dev.	Minimum	Maximum
Endogenous variables				
$health_t$	50.07	9.39	11.73	67.24
$employed_t$	0.76	0.43	0	1
$medical_t$	507	1130	0	37128
Variables appearing in all the equations				
age	42.27	11.95	19	65
sex	0.46	0.50	0	1
$race$	0.81	0.39	0	1
mar	0.63	0.48	0	1
edu	0.23	0.42	0	1
Additional variables appearing in the health equation				
$smoke$	0.23	0.42	0	1
$phyact$	0.55	0.50	0	1
$seatbelt$	0.77	0.42	0	1
$health_{t-1}$	50.11	9.41	13.21	67.13
Additional variables appearing in the labor force participation equation				
$numchi$	1.08	1.23	0	12
$child5$	0.20	0.40	0	1
$employed_{t-1}$	0.75	0.43	0	1
$spousein$	20372	28647	0	280777
Additional variables appearing in the medical expenditure equation				
hmo	0.38	0.49	0	1
$gatekeeper$	0.08	0.27	0	1
ppo	0.17	0.38	0	1
$spousein$	20372	28647	0	280777

1.5 Estimation results

In this section, the results for the endogenous variables are first presented in detail and then the results for exogenous variables of different equations are described briefly. Tables 1.4 and 1.5 present the estimation results for all the equations.

Overall, a significant impact of health on labor force participation is estimated. The positive sign indicates that, other things equal, better health increases the probability of labor force participation by 9.26%. The feedback effect of labor force participation on health is significant positive too. Else equal, the employed individuals are healthier than the unemployed individuals even after controlling for the last period health. This is probably due to the emotional stress of the unemployed individuals. We can see that healthier individuals spend less on their medical care and after controlling for the same health level, the employed individuals spend less than the unemployed which is unexpected but the result is not significant.

The most unexpected result is the impact of medical spending on health in the health equation. Even after correcting the endogeneity problem by two step procedure, the sign on medical spending is still significant negative. This result is really telling us something else. Maybe the way we are interpreting the result is not that appropriate. It is not saying that more medical spending lead to less health. It is just because we don't observe the health level of individuals if they don't get treated. After they get treated and spend a lot of money on their medical care, the health levels are highly possible not as high as before. That is why we observe a significant negative sign there. But the medical spending is contributing to the difference between the unobserved untreated health and the current health.

Looking at the estimation for the health equation, other than sex, race and seat belt are not significant, all the other results are quite as expected, the health decreases as individuals get older. Married and more educated ones are healthier than otherwise. Smoking is bad for health and excising regularly has a significant positive impact on health. Individuals who have preexisting conditions are far less healthy than those who don't.

For the labor force participation equation, in order to interpret the results and compare to the OLS results(Appendix 1), the marginal effect of each variable is calculated. The results are just as expected according to the existing literature. The probability of labor participation increases with age but at a reducing speed since age square is negative. Married and more educated individuals have higher probability of entering the labor force everything else being equal. If the individual has children under 5 years old or has a spouse with higher income, then this individual is more likely to exit the labor force. If the individual is employed last period, then it is highly possible for this individual to still remain in the labor force. There are also four cross terms too.

As for the medical expenditure equation, the results are as expected too. Older ones and whites spend more. Female spend a lot more than male. It is probably due to pregnancy or cosmetic reasons. More educated ones and

Table 1.4: Estimation results:

Variables in the Equations	Coefficients	Standard error	Marginal Effects
Health equation			
$cons$	2.6732	0.2096	
$employed_t$	0.2153	0.0249	
$medical_t/1000$	-.5434	0.2063	
$age/10$	-.0917	0.0120	
sex	0.0295*	0.0202	
$race$	0.0290*	0.0224	
mar	0.0512	0.0160	
edu	0.1003	0.0216	
$smoke$	-.0650	0.0183	
$phyact$	0.0946	0.0150	
$seatbelt$	0.0275*	0.0176	
$health_{t-1}/10$	0.4845	0.0387	
$health_{t-1}/10 medical_t/1000$	0.1095	0.0480	
R^2		0.4687	

the ones whose spouses have higher incomes spend more. Most importantly, we can see that individuals who have hmo health insurance plans spend less than others.

1.6 Conclusion

This paper has examined the relationship between health, labor force participation and medical spending using data form MEPS panel 6. The potential endogeneity problems of these three factors are addressed by estimating the health equation, labor force participation equation and medical spending

Table 1.5: Estimation results(Continued):

Variables in the Equations	Coefficients	Standard error	Marginal Effects
Labor force participation equation			
cons	-8.1171	0.6819	
$health_t$	0.7177	0.0729	0.0926
age	1.8836	0.2654	0.2431
sex	0.0841*	0.1144	0.0086
race	0.0987*	0.1078	0.0104
mar	0.4429	0.1328	0.0476
edu	0.3206	0.1162	0.0308
$age2$	-.2423	0.0306	-.0313
numchi	-.0946*	0.0780	-.0122
$child5$	-.5008	0.2563	-.0577
$employed_{t-1}$	4.2207	0.0862	0.7141
spousein	-.0098	0.0017	-.0013
sexnchi	0.2861	0.0868	0.0290
$sexchi5$	0.3200	0.2795	0.0297
marnchi	-.1698*	0.0902	-.0166
$marchi5$	0.1105*	0.2815	0.0110
R^2		0.6554	
Medical expenditure equation			
cons	1.4103	0.1290	
$employed_t$	-.0570*	0.0400	
$health_t$	-.2916	0.0227	
age	0.1248	0.0109	
sex	-.1562	0.0240	
race	0.1935	0.0298	
mar	-.0747	0.0289	
edu	0.2183	0.0291	
hmo	-.1468	0.0264	
gatekeeper	-.0490*	0.0447	
ppo	0.0507*	0.0340	
spousein	0.0017	0.0005	
R^2		0.0891	
No of observations		8896	

* not significant at 5%

equation in a simultaneous equations model by two step procedure.

Overall, a significant impact of health on labor force participation is estimated. The feedback effect of labor force participation on health is significant positive too. This is probably due to the emotional stress of the unemployed individuals. We can see that healthier individuals spend less on their medical care and after controlling for the same health level, the employed individuals spend less than the unemployed which is unexpected but the result is not significant.

There are two potential issues for this kind of reduced form regression: interpretation and identification.

How do we supposed to interpret the negative sign on medical spending in the health equation? It is not saying that more medical spending lead to less health. It is just because we don't observe the health level of individuals if they don't get treated. After they get treated and spend a lot of money on their medical care, the health levels are highly possible not as high as before. That is why we observe a significant negative sign there. But the medical spending is contributing to the difference between the unobserved untreated health and the current health.

In order to make this simultaneous equations model work and identify the model, there have to be some reasonable good exogenous excluded variables. Are those exogenous excluded variables mentioned in Section 1.4.2 good enough? There is always plenty of criticism against those variables being used as instruments. But because of the data limitation, those variables are the best the author could find in this data set.

Because of the issues mentioned above, the author calls for a structural model of this kind of relationship which can model individuals preference and behavior, technology and market structure.

Chapter 2

A Medical Expenditure Puzzle

What does your medical care do to your health? Researchers often find a significant negative coefficient on medical expenditure in reduced form health production regressions. The puzzling result motivates this dynamic heterogeneous agent model to study the relationships between health status, medical expenditure and employment. In this model, individuals value both consumption and health and they are heterogeneous in their health levels and employment status. Individuals choose how much to spend on medical care and consumption. Health can be accumulated from investment in medical care and this increases both job opportunities and quality of life. The structural parameters are estimated by an indirect inference procedure which matches the simulated regression coefficients to the data regression coefficients from the Medical Expenditure Panel Survey. This paper finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure impacts the health in the positive way. This paper also explains the basis for this result.

2.1 Introduction

What does your medical care do to your health? People would respond very naturally that of course it will improve health. Else, there is just no point of spending any money on medical care. However, researchers often get a significant negative coefficient on medical expenditure in reduced form health production regressions. Researchers attribute this to the correlation between medical expenditure and something in the error term which causes a downward bias. What is that something in the error term and why can that something cause the downward bias? This puzzling result motivates this paper to investigate the role of medical expenditure in improving health. This paper explains why we have this puzzling result in reduce form regressions while at the same time medical spending impacts the health in the positive way.

Medical expenditure has been suggested to be viewed as one form of human capital decades ago (Mushkin 1962, pp. 129-49; Becker 1964, pp. 33-36; Fuchs 1966, pp. 90-91).[1] Grossman (1972) is the first to construct a life cycle model of the demand for health capital itself. In his model, health is viewed as a capital stock which yields an output of "healthy days". Individuals may invest in health by combining time (e.g., for doctor's visits) with purchased inputs (e.g., medical services). The incentive for investing in health is that by increasing the health stock the individual increases the amount of time available for earning income or for producing consumption goods. One prediction

[1] Quoted from Grossman 1972.

of the model is that if the health depreciation rate increases with age, at least after some point in the life cycle, then the quantity of health demanded would decline over the life cycle. This prediction relies crucially on the assumption that consumers fully anticipate intertemporal variations in depreciation rates and therefore, know their age of death with certainty. At the same time, provided the elasticity of the marginal efficiency of capital schedule were less than unity, expenditures on medical care would rise with age. As a result, under certain conditions, health demanded and medical care demanded could go in the opposite direction.

Grossman (1972b) then runs a reduced form regression on health production function and finds that the coefficient on medical expenditure is significant negative. He explains this as a result of the correlation between the medical expenditure and the error term (health depreciation rate). The correlation causes a downward bias and hence, the coefficient could be negative. The explanation is not quite satisfactory because age is already one of the independent variables in that regression. However, according to Grossman (1972a), health depreciation rate only varies with age. Furthermore, even after controlling age and initial health, the coefficient on medical expenditure in the health production function is still negative based on the empirical work in this paper.

In order to correct the wrong sign on medical expenditure, a lot of labor and health economists have devoted enormous amount of efforts to find the right data set and right instruments to a certain group of individuals.

Rosenzweig and Schultz (1983, 1988, 1991) and Grossman and Joyce(1990) consider the effects of some health inputs such as mother's prenatal care and smoking behavior on baby's health. They use baby's weight as a measure of baby's health and use income, education and price of medical care service as instruments for medical care. They get the consistent estimates though the validity of baby's weight as an instrument for baby's health has been questioned.

Instead of spending all the efforts on searching for the good instruments and trying to correct the sign as the most researchers do, this paper studies the role of medical expenditure in improving health in a heterogeneous agent model. In this model, individuals value both consumption and health and they are heterogeneous in their health levels and employment status. Individuals choose how much to spend on medical care and consumption. Health can be accumulated from investment in medical care and this increases both job opportunities and quality of life. The structural parameters are estimated by an indirect inference procedure which matches the simulated regression coefficients to the data regression coefficients from the Medical Expenditure Panel Survey. This paper finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical care impacts the health in the positive way.

The reason why this paper has this result is because there are two roles played by medical care in improving health. One is "curing" role and the other is "improvement" role. The "curing" role captures the medical care of recovery from a sudden drop of health level(i.e. acute disease or accidents) and the "improvement" role captures the care from individuals' investment motive. In a cross section reduced form estimation of health production function, if the effect of "curing" dominates the effect of "improvement", the coefficient on medical expenditure will be negative. Hence, the gain of medical care is often hidden in reduced form regressions. However, with the structural model in this paper, people can see clearly how medical care improve health in the health evolution equation.

There are papers which incorporate medical expenditure risks into explaining individuals' savings, insurance choices, bankruptcy filing, retirement decision and probably more. The saving literature includes Kotlikoff (1989), Hubbard, Skinner and Zeldes (1994, 1995), Palumbo (1999) and De Nardi, French and Jones (2005). Kotlikoff (1989) examines the effects of different financing mechanisms for random health expenditures on macroeconomic saving rates. Hubbard, Skinner and Zeldes build a life-cycle model that incorporates uncertainty regarding annual earnings, medical expenses and longevity to explain the distribution of wealth holdings in the U.S. (their 1994 paper) and to study the consequences of a resource-tested Medicaid program for saving decisions by low- and middle-income families (their 1995 paper). Palumbo (1999) and De Nardi, French and Jones (2005) incorporate medical expenditure risks to study savings of the elderly. Jeske and Kitao (2005) work on the income tax policy with health insurance choice in a general equilibrium framework. Chatterjee, Corbae, Nakajima and Rios-Rull (2005) takes into account the unpaid heath-care bills for the bankruptcy filing decisions. Rust and Phelan (1997) and Blau and Gilleskie (2005) look into the association between retiree health insurance and employment of older individuals with medical expenditure risks in the model.

However, all the research mentioned above take medical expenditure risks as exogenously given instead of as endogenous choice. As a result, medical expenditure can only take the "curing" role to restore the health to its original level in the papers mentioned above. This paper is among the first to endogenize both medical expenditure choices and individuals' health levels. Hall and Jones (2006) argues that the rise in health spending over the years is a rational response to the growth of income since the elasticity of health spending to income is above one. That paper concentrates on the role of medical expenditure to improve health at aggregate level (i.e. solving a social planner's problem) and therefore, health in that model is the inverse of mortality rate. The rise of medical care investment of the nation decreases mortality rate and increases average life span which drive the elasticity of total health spending to total income above one. My paper investigates the role of medical spending to improve individuals' health levels and in turn, how health impacts individuals' job opportunities and quality of life. It is rich enough to incorporate social insurance and at the same time, provides room for policy experiments.

This paper has the following contributions. First, it looks deeply into the role of medical spending in improving individuals' health in a structural model and therefore, it can explain the puzzling result in reduced form regressions. Second, it provides a quantitative model in which individuals value both their consumption and health and at the same time, health can be accumulated through medical care investment. Third, it pins down the weight between consumption and health in individuals' preference and the parameters in the structural health production function. Last, by estimating the structural parameters with indirect inference procedure, it bridges reduced form regressions with the structural estimation.

The paper is arranged as follows: Section 2 outlines the theoretical model. Section 3 describes the model parameterization strategy and the data. Section 4 presents the estimation results and the analysis. Section 5 concludes the paper and discusses the future work.

2.2 The Theoretical Model

This paper considers a dynamic quantitative model with heterogeneous agents. The various components of the theoretical model are described below.

2.2.1 Preference and Health Production Function

The economy is populated by a large number of individuals who are ex ante heterogeneous with respect to their health status and employment status. The agents in this economy are infinite-lived and maximize the following expected value of their discounted utility:

$$E\sum_{t=0}^{\infty} \beta^t U(c_t, h_t) \tag{2.1}$$

where c_t is consumption, h_t is health stock, β is the discount factor and $0 < \beta < 1$, $U(\cdot, \cdot)$ is the momentary utility function.

In this simple model, agents can't save. Savings will be incorporated into one of the possible extensions of this paper. Their budget constraint is given by

$$c_t = y_t^d - m_t \tag{2.2}$$

where c_t is the consumption in the current period, y_t^d is the disposable income in the current period and m_t is the individual's out-of-pocket medical care expense in the current period.

Agents' health stock evolves according to the following equation:

$$h_{t+1} = \phi_t(1 - \delta)h_t + a(m_t)^b \tag{2.3}$$

where h_t is the health stock at the beginning of the current period, δ is the health depreciation rate, ϕ_t is the health shock at the beginning of the current period, a and b are parameters and h_{t+1} is the health stock at the beginning of the next period. For the ease of later analysis , the medical expenses incurred to restore the health back up to its starting level in the period because of a bad health shock are referred as medical expenses for "curing" purpose and other medical expenses are referred as medical expenses for "investment" purpose.

The health shock ϕ_t is an i.i.d. shock which takes two values $\phi \in \Phi = \{\phi_g, \phi_b\}$[2].

[2] ϕ_g is no shock in this model and ϕ_b is bad shock, i.e. a car accident or a broken leg. $\phi_g = 1$ and $\phi_b = 1 - z$.

2.2.2 Job Opportunities and Markov Transition Function

In each period of their lives, agents face a stochastic employment opportunity. Let s denote the employment state of an individual. If $s = e$, the agent is employed and if $s = u$, the agent is unemployed. Conditional on agents' employment status last period and health stock at the beginning of this period, the employment probabilities in this period are denoted by $\pi(e'|e, h')$, $\pi(u'|e, h')$, $\pi(u'|u, h')$ and $\pi(e'|u, h')$. $\pi(e'|e, h')$ and $\pi(e'|u, h')$ are estimated by Nadaraya Watson nonparametric regression[3] from the MEPS data (Figure 1). $\pi(u'|e, h')$ and $\pi(u'|u, h')$ are just $1 - \pi(e'|e, h')$ and $1 - \pi(e'|u, h')$ respectively. Given a particular value of h', employment status transitions follow a two state markov process.

Figure 3.1 shows the employment probabilities this period conditional on whether the individual is employed (top) or unemployed (bottom) last period and their health status. The vertical axis is the estimated employment probabilities and the horizontal axis is the actual health stock divided by 100. The lowest health level is 11.73 and the highest level is 67.24 in the MEPS data (see Table 2). The probabilities are continuous in h'. From the graphs, we can see that everything else being equal, a healthier agent has a higher chance of getting the job and the agent who had a job last period has a better chance of being employed this period.

In MEPS, employment status is surveyed every half a year but medical

[3]Adrian Pagan and Aman Ullah (1999), Nonparametric Economics, Cambridge University Press.

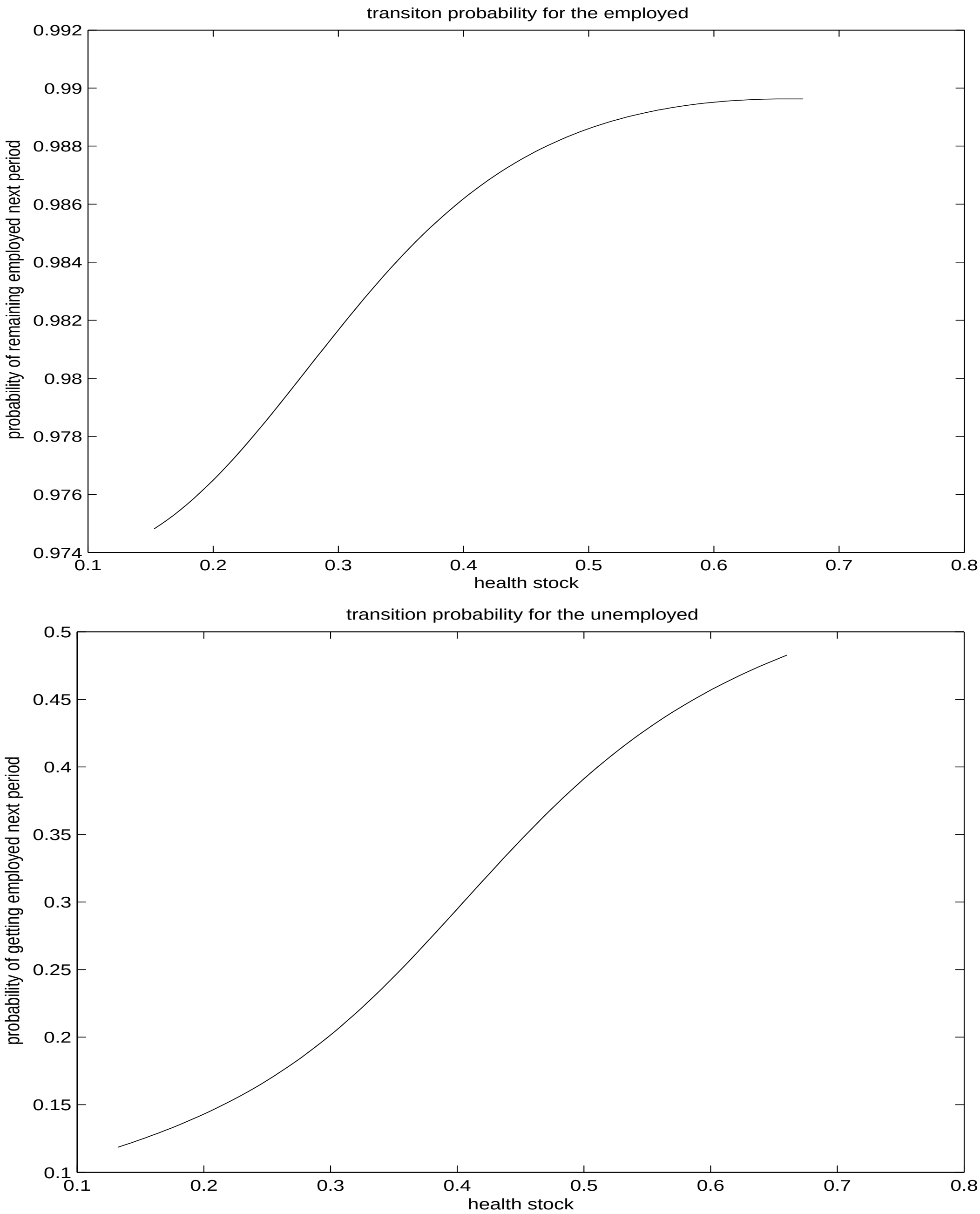

Figure 2.1: The employment transition probabilities for employed and unemployed individuals

expenditure is the expenditure accumulated within one year. In order to create yearly data, the individual is assumed to be employed as long as he is employed in either half of the year. This will result in an underestimate of the impact of health status on individual's employment status and increase the persistence of employment status in the top graph of Figure 3.1. It would be better if we have monthly or quarterly data in order to estimate these transition functions more accurately. However, it won't have a big impact on the robustness of the structural parameter estimates later on since the same sort of bias applies to the reduced form data regression.

If the agent is employed, his disposable income is $y - \tau$, where y is his income and τ is the income tax he has to pay. If the agent is unemployed, his disposable income is just his unemployment insurance from the government θy, where θ is the replacement ratio of unemployment insurance.

Let η denotes the employment status of the agent. If the agent is employed, $\eta = 1$, otherwise $\eta = 1$.

2.2.3 Timing and Defining Maximization Problem

The timing of the model is the following: The agent enters period t with a health stock of h_t and then his employment status is revealed. Then a health shock ϕ_t is realized. After observing his health shock, the agent chooses m_t either to recover from a bad shock and/or improve his health. Then he enters period t+1 with a health stock of h_{t+1} which evolves according to equation (3.3).

The maximization problem can be written as a dynamic programming problem. Note that the state variable are health stock h, employment opportunity s and health shock ϕ. The dynamic programming problem is:

$$\begin{aligned} V(\phi, e, h) &= \max_m u[y(1-\tau) - m, h] + \beta \sum_{\phi'} \chi(\phi'|\phi) \sum_{s'} \pi(s'|e, h') V(\phi', s', h') \\ V(\phi, u, h) &= \max_m u[\theta y - m, h] + \beta \sum_{\phi'} \chi(\phi'|\phi) \sum_{s'} \pi(s'|u, h') V(\phi', s', h') \end{aligned}$$

subject to

$$m \geq 0 \tag{2.4}$$

2.3 Parameterizations

This section first describes how those parameter values in the theoretical model are chosen and what parameters need to be estimated. Then it presents the steps of this indirect inference estimation strategy. The data used in the estimation is described in the third subsection. A descriptive model which is the basis for this indirect inference procedure is then formulated at the end of this section.

2.3.1 Parameters in the Structural Model

The utility function used in the computation has the following form:

$$U(c, h) = \frac{(c^{1-\sigma} h^{\sigma})^{1-\rho} - 1}{1-\rho} \tag{2.5}$$

where $0 \leq \sigma < 1$ and $\rho > 0$ guarantee that $u_c > 0$, $u_h > 0$, $u_{cc} < 0$ and $u_{hh} < 0$ but the sign of u_{ch} depends on whether ρ is greater or less than 1.

The time period in the model is one year. The health stock is set between 0 and 1 to match the data. β is set to 0.95. δ is the health depreciation rate in the health evolution equation. δ is set to 0.01 based on the following calculation. The lowest health level in the data is 11.73 and the highest is 67.24. This paper assumes that human beings can live up to 150 years if without any accident or illness. Then, $67.24(1-\delta)^{150} = 11.73$. This gives us the value of δ. If this paper assumes that human beings can live up to 100 years without any accident or illness, δ is calculated to be 0.017. This does not affect the main results of this paper.

The wage rate y is normalized to 1. This paper takes policy parameters θ and τ as exogenous. θ is set to 0.25 for the following reason. The U.I. replacement ratio in the United States is about 0.5 which usually lasts for six months. Since the unemployment duration in this model is assumed to be 1 year so the U.I. replacement ratio is set to be 0.25 in this paper and τ is set to 0.15 which is the average federal tax rate in the United States.

The parameters need to be estimated by Indirect Inference are $\Theta = \{\sigma \ \rho \ a \ b \ z\}$. σ and ρ are preference parameters. a and b are parameters in the health stock evolution equation. z is the health shock parameter.

2.3.2 Indirect Inference

The structural model parameters are estimated using the method of indirect inference. For arbitrary values of the vector of parameters Θ, the dynamic programming problem is solved and policy functions are generated.

Using these policy functions, the decision rule is simulated to create a simulated version of the data to match. One then chooses a descriptive statistical model that provides a rich description of the patterns of covariation in the data. Such a descriptive model can be estimated on both the simulated data from the structural model, and on the actual observed data. This then gives us two sets of coefficients to match, $\Psi^s(\Theta)$ and Ψ^d.

The estimate $\hat{\Theta}$ is pinned down by minimizing the weighted distance between the actual and simulated coefficients from the descriptive models. Formally, it solves

$$\pounds(\Theta) = \min_{\Theta}[\Psi^d - \Psi^s(\Theta)]' W[\Psi^d - \Psi^s(\Theta)] \tag{2.6}$$

where W is a weighting matrix. The method of indirect inference will generate a consistent estimate of θ. The weighting matrix,W, is constructed as the inverse of the variance-covariance matrix of the coefficients estimated from the MEPS.

Since the $\Psi^s(\Theta)$ function is not analytically tractable, the minimization is performed using numerical techniques. A simulated annealing algorithm is used to perform the optimization in order to obtain the global minimum in parameter space independent of starting values.

2.3.3 The Data

The data used for this paper come from the Household Component of the Medical Expenditure Panel Survey. The MEPS HC is a nationally representative survey of the U.S. civilian noninstitutionalized population. It collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a -year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures. MEPS HC panel 6 covers two years' data (2001 and 2002) for 21,959 individuals. Since this paper's main goal is to explore the relationships for the working age individuals, the sample of individuals whose ages are either below 19 or over 64 is dropped. Students are out of sample too. After cleaning the sample by deleting the observations with missing information, 7515 data points are left.

2.3.4 Descriptive Model

The descriptive model consists of two linear equations which are extensively estimated in the literature of health economics.[4] These equations are health equation and medical expenditure equation. They take the following forms:

$$\begin{aligned} h_{t+1} &= \alpha_1 m_t + \alpha_2 h_t + \alpha_3 h_t m_t + X_1 \alpha_4 + \varepsilon_{1t} \\ m_t &= \gamma_1 h_t + \gamma_2 s_t + X_2 \gamma_3 + \varepsilon_{2t} \end{aligned} \tag{2.7}$$

where h is individual's health, s is individual's employment status and m is individual's out-of-pocket medical expenditure.[5] X_1 and X_2 are control variables in these equations. Since these control variables are not modeled in the structural model, it is assumed that the agents in the structural model are homogenous in them. The MEPS coefficients which are going be matched by the simulated coefficients are $\{\alpha_1\ \alpha_2\ \alpha_3\ \gamma_1\ \gamma_2\}$. Tables 2.1 and 2.2 provide the definitions and the summary statistics of all the variables in these equations.

In the health equation, medical expenditure is expected to have a positive effect on health but the coefficient in the OLS regression is usually negative. Last period health is expected to have a positive effect on this period health.

---

[4]Currie and Madrian (1999) has a detailed review of them. Stratmann (1999) estimates the effect of doctor visits on work day loss using the types of health insurance as instruments for doctor visits.

[5]Here, out-of-pocket medical expenditure is in the form of the percentage of income in order to match the simulated data from the structural model. In order to use the medical expenditure share, the individuals without any income or with the medical expenditure share greater than 1 are out of sample.

In the medical expenditure equation, conditional on the same employed status, healthier individuals are expected to spend less on their medical bill and therefore, the coefficients is expected to be negative. Conditional on the same health level, the employed individuals are expected to spend more on their medical care and therefore the coefficient is expected to be positive but in the reduced form regressions, the coefficient is significantly negative.

The coefficients on the control variables in the two equations all have their expected signs. These two equations are estimated separately by OLS regression. The results are reported in Table 2.3. According to the OLS results,

$$\Psi^d = \{-0.4249\ 0.5325\ 0.5508\ -0.0915\ -0.0440\}.$$

2.4 Estimation Results

The estimation procedure described in the previous section gives us the following results of the structural parameters and the coefficients from the simulated data:

From Table 2.5, we can see that the simulated coefficients match the MEPS coefficients quite well. Interestingly, the simulated α_1, the coefficient of medical expenditure in the health equation, is negative even though in the health evolution equation of the structural model, medical expenditure only

Table 2.1: Variable definitions:

Dependent variables	
$health_{t+1}$	physical component summary index* at the end of period t, 0 is the lowest heath level, 100 is the highest
$medical_t$	total amount of out-of-pocket medical expenditure
Independent Variables appearing in all the equations	
age	Individual's actual age
sex	1 if male, 0 otherwise
race	1 if white, 0 otherwise
mar	1 if married, 0 otherwise
Additional independent variables appearing in the health equation	
$health_t$	physical component summary index at the end of period t-1, 0 is the lowest heath level, 100 is the highest
Additional independent variables appearing in the medical expenditure equation	
$employed_t$	employment status at period t: 1 if in labor force, 0 otherwise
inscov	whether the individual has health insurance coverage: 1 if yes, 0 otherwise

*Refer Appendix 1 for the description of health measure used in this paper.

Table 2.2: Descriptive statistics of the sample:

Variables	Sample Mean	Standard dev.	Minimum	Maximum
Dependent variables				
$health_{t+1}/100$	0.51	0.08	0.1173	0.6724
$medical_t/income_t$	0.02	0.06	0	0.9983
Independent Variables appearing in all the equations				
$age/100$	0.41	0.11	0.19	0.64
sex	0.49	0.50	0	1
$race$	0.81	0.39	0	1
mar	0.62	0.49	0	1
Additional independent variables appearing in the health equation				
$health_t/100$	0.51	0.09	0.1321	0.6713
Additional independent variables appearing in the medical expenditure equation				
$employed_t$	0.94	0.23	0	1
$inscov$	0.85	0.36	0	1

Table 2.3: Estimation results of OLS estimation

Variables in The Equations	Coefficients	Standard Error
Health equation		
cons	0.2646	0.0066
$medical_t/income_t$	-.4249	0.053
$health_t/100$	0.5325	0.0103
$(health_t/100)(medical_t/income_t)$	0.5508	0.1139
$age/100$	-.0735	0.0070
sex	0.0047	0.0016
race	0.0028*	0.0020
mar	0.0042	0.0016
R^2	0.3962	
Medical expenditure equation		
cons	0.1043	0.0061
$health_t/100$	-.0915	0.0087
$employed_t$	-.0440	0.0032
$age/100$	0.0544	0.0064
sex	-.0146	0.0014
race	0.0053	0.0018
mar	-.0039	0.0015
inscov	-.0125	0.0020
R^2	0.0912	
No of observations	7515	

*the coefficient is not significant at 5 percent significant level.

Table 2.4: Structural Parameter Estimates:

	σ	ρ	a	b	z
Estimated Value	0.0518	1.5321	0.5702	0.3746	0.3333
Standard Error	0.0007	0.0040	0.0020	0.0005	0.0006

Table 2.5: Coefficients and Moments from the MEPS and Simulated Data:

	α_1	α_2	α_3	γ_1	γ_2	$\pounds(\Theta)$
MEPS	-0.4249	0.5325	0.5508	-0.0915	-0.044	
Simulated	-0.4739	0.5208	0.606	-0.0251	-0.0652	45

impacts the health in the positive way because parameters a and b are both positive.

Why is the estimated α_1 negative? In the simulated reduced form health production function regression, health level at period t is controlled for. Hence, I only need to analyze the individuals with the same health level at period t. The medical expenses of these individuals differ because of the realization of health shocks and employment status. There are two forces which impact the sign of the coefficient on medical expenditure in the simulated reduced form health production regression.

On the one hand, for the individuals with the same starting health and the same employment status, some get no health shock and some get a bad health shock. Furthermore, health will depreciate even if no health shock is realized. The individuals with no health shock might spend money on medical care for investment purpose. The individuals with bad health shock have to spend money on medical care to cure themselves first and then invest. As a result, even though the individuals with the bad health shock might spend much more money than the individuals with no adverse health shock, the health outcome could be much lower depending both on the size of the bad shock and the parameters in health evolution function. This gives us the result of more medical spending but with worse health outcome. The bad health shock is the key element to this negative coefficient on medical expenditure.

On the other hand, individuals in this model differ from each other because their employment status and therefore, their income level too. The income dimension coincides with the employment dimension in this model since the heterogeneity of income solely comes from the heterogeneity of employment status. Let us compare the employed individuals with no health shock and the unemployed individuals with bad health shock with the same starting health level. Unemployed individuals have much tighter constraints than employed individuals and they have to balance their resources between consumption and medical care. If the starting health level is low, even if the employed have a bad health shock, their medical spending might be less than the medical spending of the employed. That will give us the result of more medical spending with better health outcome.

As a result, depending on the structural parameters and composition of the sample, the coefficient on medical expenditure in the reduced form health production regression could either be positive or negative. The reason why we have the negative coefficient is because the medical expenditure for "curing" purpose dominates the medical expenditure for "investment" purpose.

The analysis above is tested by the following experiment. The descriptive model regressions are redone controlling for the health shock in the health production function. If the analysis above is true, then the coefficient on medical expenditure in the health production function regression should be positive after controlling for the health shock while the coefficient on health shock should be negative. The experiment results are presented in Table 2.6.

Table 2.6: Coefficients from the MEPS Data and Coefficients from the Simulated Data without and with Controlling for Health Shocks:

	α_1	α_2	α_3	α_4
MEPS	-0.4249	0.5325	0.5508	NA
Simulated No Control	-0.4739	0.5208	0.606	NA
Simulated Control	0.3182	0.5458	0.1478	-0.1446

As described in last section, α_1, α_2 and α_3 are the coefficients on medical expenditure, previous health and the cross term of these two respectively in the health equation. *MEPS* and *Simulated No Control* present the estimation results of the above two regressions by using the MEPS data and the simulated data respectively. *Simulated Control* presents the estimation results of the above two regressions but with *healthshock* as one of the independent variable in the health equation by using the simulated data. Indeed, the coefficient on medical expenditure in the health equation becomes positive in the above experiment. At the same time, the coefficient on health shock,

α_4, in the health equation is negative. The reason why researchers often get a significantly negative coefficient on medical expenditure in the OLS health production regression is because they don't observe health shocks in the real data.

The fact that the simulated coefficients match the data coefficients quite well reflects that the theoretical model is reasonably specified and there exists a set of structural parameters which make simulated individuals' choices mimic individuals' choices in reality.

From a statistical perspective, the model is rejected since the reported values of $\pounds(\Theta)$ are still high compare to the cut off value. However, in this setting, this reflects the fact that the coefficients are calculated from a very large panel data set, implying very small standard deviations of the coefficients (and a very large W). Given how precisely the micro coefficients are estimated from the actual data, virtually any model would be formally rejected with even very modest deviations of the simulated coefficients from the actual coefficients. As we have emphasized above, the fit of the model in the last line of Table 2.5 is actually quite good in terms of matching the data coefficients on both a qualitative and quantitative basis.

2.5 Conclusion

This paper is motivated by a "medical expenditure puzzle". Instead of trying to correct the sign of the coefficient on medical expenditure in the reduced form health production function, this paper studies individuals' choice of medical expenditure and how their medical care affect their health and in turn their job opportunities and quality of life in a dynamic heterogeneous agent model. Through this study, we can see clearly how the medical care impacts individuals' health and what factors determine how much individuals are going to spend on their medical care.

The structural parameters are estimated by the method of Indirect Inference which minimizes the distance function of the MEPS coefficients and the simulated coefficients. The idea of this indirect inference procedure is that if the model is correctly specified, the coefficients from the simulated data regressions should have the same direction and degree of bias as the coefficients from the MEPS data regressions. This is the first study to estimate the weight and risk averse parameters in a CRRA utility function with both consumption and health and at the same time, estimate the health production function.

This paper finds that the simulated coefficient of medical expenditure in the health equation is negative even though in the health evolution equation of the structural model, medical expenditure impacts health in the positive way. The reason why we have the negative coefficient is because the medical expenditure for "curing" purpose dominates the medical expenditure for "investment" purpose. This study suggests that reduced form regression results sometimes give us the wrong information without knowing the structural model behind it. Furthermore, it is much easier to see agents' behavior and the causalities of key variables in the structural model. The structural model also provides room for policy experiments which will be the natural extensions of this paper.

Chapter 3

Measuring the Welfare Cost of Mandatory Employer-provided Health Insurance

Empirically, 83 percent of the 45 million uninsured in the United States are in working families. As a result, mandatory employer-provided health insurance has been one of the hotly debated policies in order to reduce the portion of the uninsured. However, whether this mandate is economically justified remains an open question.

There are two possible effects of this mandate. On the one hand, employees might choose to be uninsured if they are relatively healthy and hence do not value health insurance. After the mandate, these employees could be worse off. On the other hand, uninsured employees tend to receive less therapeutic care when facing accidents or diseases. This might drive them out of labor force which imposes a cost to society in terms of both a loss of tax income and increased burden to welfare program. In that case, society could become better off because of this mandate.

This paper concentrates on measuring the welfare cost of this mandate in a dynamic heterogeneous agent model. In this model, individuals value both consumption and health and they are heterogeneous in their health levels and employment status. Individuals choose whether to have health insurance or not and how much to spend on medical care and consumption. Health can be accumulated from investment in medical care and this increases both job opportunities and quality of life. The structural parameters are estimated by an indirect inference procedure which matches the simulated regression coefficients to the data regression coefficients from the Medical Expenditure Panel Survey. With the structural parameters being estimated, this paper then compares the welfare of the working-aged individuals before and after the mandate. It finds a welfare cost of 0.7 percent of GDP.

3.1 Introduction

There are about 45 million uninsured individuals in the United States. 83 percent of the uninsured are in a family with a worker (Fronstin, 2004). As a result, the mandatory employer-provided health insurance has been one of the hotly debated policies in order to reduce the portion of the uninsured.

The policy has been proposed by Clinton administration but in vain in early 1990s. The objective of ensuring adequate health care for everyone in society is admirable, and the proposal for mandatory, employer-provided insurance is well-intentioned. However, it is quite possible that federally-mandated but privately-financed health insurance would actually harm its intended beneficiaries. People could become uninsured because they don't value health insurance and hence choose to work in the firm without health insurance but with higher salary or choose to decline the health insurance coverage when offered in exchange for a higher salary. After the mandate, firms will eventually transfer the cost of health insurance provision back to their workers to make even. Indeed, Gruber and Krueger (1990) finds that 86.5 percent of the cost of mandated health insurance may be shifted to the worker, primarily in the form of lower money wages, using evidence from state workers' compensation insurance.

On the other hand, uninsured employees tend to receive less therapeutic care when facing accidents or diseases. This might drive them out of labor force which imposes a cost to society in terms of both a loss of tax income and increased burden to welfare program. In that case, society could become better off because of this mandate.

This paper concentrates on measuring the cost of mandatory employer-provided health insurance policy by comparing the welfare of the working aged individuals with and without this policy. It finds that there is a welfare cost 0.7 percent of GDP associated with this mandate. This paper allows the individuals to choose their own medical expenses to improve their health status and in turn increase their job opportunities when facing the unexpected health shocks in a dynamic heterogenous agent model. In this model, they also have health insurance choice according to their own characteristics. The structural parameters in the model are estimated by matching the model's implications with individual observations from the Medical Expenditure Panel Survey (MEPS) as part of a minimum distance estimation routine.

This paper is the first to endoginize individuals' health status and therefore health insurance and medical expense choices in a dynamic heterogeneous agent model. This paper bridges the reduced form regression with structural estimation by Indirect Inference procedure. It successfully estimates the preference parameters of the structural model with endoginized health and the parameters of the health production function which usually has the wrong sign problem in the reduced form regression (Han 2005).

The paper is arranged as follows: Section 2 provides some history of the employer-provided health insurance. Section 3 describes the data. Section 4 outlines the theoretical model. Section 5 describes the model parameterization strategy and presents the estimation results. Section 6 presents the policy experiment. Section 7 concludes the paper and discusses the future work.

3.2 History of the Employer Provision of Health Insurance

During the late 1800s, companies in the railroad, mining, lumber, and other industries began hiring company doctors (Institute of Medicine, 1993). Employers in these industries provided company doctors funded by deductions from workers' wages. The employees in these industries often worked in isolated areas where replacement workers were difficult to find, and the company self-interest in returning injured or sick workers to full health in such circumstances is self-evident.

During the World War II, many employers began to offer health benefits to get around the wage controls set by the National War Labor Board in order to attract scarce workers. As a result, the number of persons with employment-based health benefits started to increase. In 1943, the National War Labor Board ruled that employer contributions to insurance did not count as wages, and, thus, did not increase taxable income. Historians often suggest that the tax-preferred status of employment-based health benefits let to the rise in its prevalence and comprehensiveness (Fronstin, 2006). More than 159 million individuals under age 65 had some form of employment-based health benefits during 2004 (Fronstin, 2005).

For the unemployed, an mix of other institutions has been developed to "fill-in-the-gaps": Medicare for those over 65 (the "retired") and the permanently disabled; Medicaid for children in lower income families and women who are on welfare; a small non-group private insurance market for the self-employed or individuals otherwise lacking insurance; and other miscellaneous programs such as university-provided health insurance for students who are no longer dependents of their parents(Currie and Madrian, 1999).

In 2003, out of 252.7 millions nonelderly Americans, the percentage covered by employment-based health benefits is 63 percent (159.2 millions). The percentage of individually purchased health insurance is 6.7 percent (17 millions). The percentage covered by public programs is 16.8 percent(42.5 millions). 17.7 percent of nonelderly Americans (44.7 millions) has no health insurance coverage. While public programs, such as Medicare and Medicaid, have been aimed at those farthest from the reach of employment-based coverage-particularly the elderly, disabled, and families with no or very low income, the problem of the uninsured is closely tied to the work place. About 83 percent of the 44.7 million uninsured are in a family with a worker (Fronstin, 2004). As a result, the mandatory employer-provided health insurance has been a hotly debated policy in order to reduce the portion of the uninsured.

3.3 The Data

The data used for this paper come from the Household Component of the Medical Expenditure Panel Survey. The MEPS HC is a nationally representative survey of the U.S. civilian noninstitutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment. The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a -year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for two calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures. MEPS HC panel 6 covers two years' data (2001 and 2002) for 21,959 individuals. Since this paper's main goal is to explore the relationships for the working age individuals, the sample of individuals whose ages are either below 19 or over 64 is dropped. Students are out of sample too. After cleaning the sample with the missing information, 7515 data points are left.

3.4 The Theoretical Model

This paper considers a dynamic quantitative model with heterogeneous agents. The various components of the theoretical model are described below.

3.4.1 Preference and Health Production Function

The economy is populated by a large number of individuals who are ex ante heterogeneous with respect to their health status and employment status. The agents in this economy are infinite-lived and maximize the following expected value of their discounted utility:

$$E\sum_{t=0}^{\infty}\beta^t U(c_t, h_t) \tag{3.1}$$

where c_t is consumption, h_t is health stock, β is the discount factor and $0 < \beta < 1$, $U(\cdot,\cdot)$ is the momentary utility function.

In this simple model, agents can't save. Savings will be incorporated into one of the possible extensions of this paper. Their budget constraint is given by

$$c_t = y_t^d - m_t^o \tag{3.2}$$

where c_t is the consumption in the current period, y_t^d is the disposable income in the current period and m_t^o is the individual's out-of-pocket medical care expense in the current period.

Agents' health stock evolves according to the following equation:

$$h_{t+1} = \phi_t(1-\delta)h_t + a(m_t)^b \tag{3.3}$$

where h_t is the health stock at the beginning of the current period, δ is the health depreciation rate, ϕ_t is the health shock at the beginning of the current period, m_t is the total medical care expense incurred in the current period, a and b are parameters and h_{t+1} is the health stock at the beginning of the next period. For the ease of later analysis , the medical expenses incurred to restore the health back up to its starting level in the period because of a bad health shock are referred as medical expenses for "curing" purpose and other medical expenses are referred as medical expenses for "investment" purpose.

The relationship between the individual's out-of-pocket medical care expense and the total medical care expense is given by the following equations. If the individual is insured

$$m_t^o = \alpha m_t$$

and if the individual is uninsured

$$m_t^o = m_t \tag{3.4}$$

where α is the co-insurance rate of the health insurance for the individual, which is just the portion of the medical bill the individual has to pay.

The health shock ϕ_t is an i.i.d. shock which takes two values $\phi \in \Phi = \{\phi_g, \phi_b\}$[1].

3.4.2 Job Opportunities and Markov Transition Function

In each period of their lives, agents face a stochastic employment opportunity. Let s denote the employment state of an individual. If $s = e$, the agent is employed and if $s = u$, the agent is unemployed. Conditional on agents' employment status last period and health stock at the beginning of this period, the employment probabilities in this period are denoted by $\pi(e'|e, h')$, $\pi(u'|e, h')$, $\pi(u'|u, h')$ and $\pi(e'|u, h')$. $\pi(e'|e, h')$ and $\pi(e'|u, h')$ are estimated by Nadaraya Watson nonparametric regression[2] from the MEPS data (Figure 1). $\pi(u'|e, h')$ and $\pi(u'|u, h')$ are just $1 - \pi(e'|e, h')$ and $1 - \pi(e'|u, h')$ respectively. Given a particular value of h', employment status transitions follow a two state markov process.

Figure 3.1 shows the employment probabilities this period conditional on whether the individual is employed (top) or unemployed (bottom) last period and their health status. The vertical axis is the estimated employment probabilities and the horizontal axis is the actual health stock divided by 100. The lowest health level is 11.73 and the highest level is 67.24 in the MEPS data (see Table 2). The probabilities are continuous in h'. From the graphs,

---

[1] ϕ_g is no shock in this model and ϕ_b is bad shock, i.e. a car accident or a broken leg. $\phi_g = 1$ and $\phi_b = 1 - z$.

[2] Adrian Pagan and Aman Ullah (1999), Nonparametric Economics, Cambridge University Press.

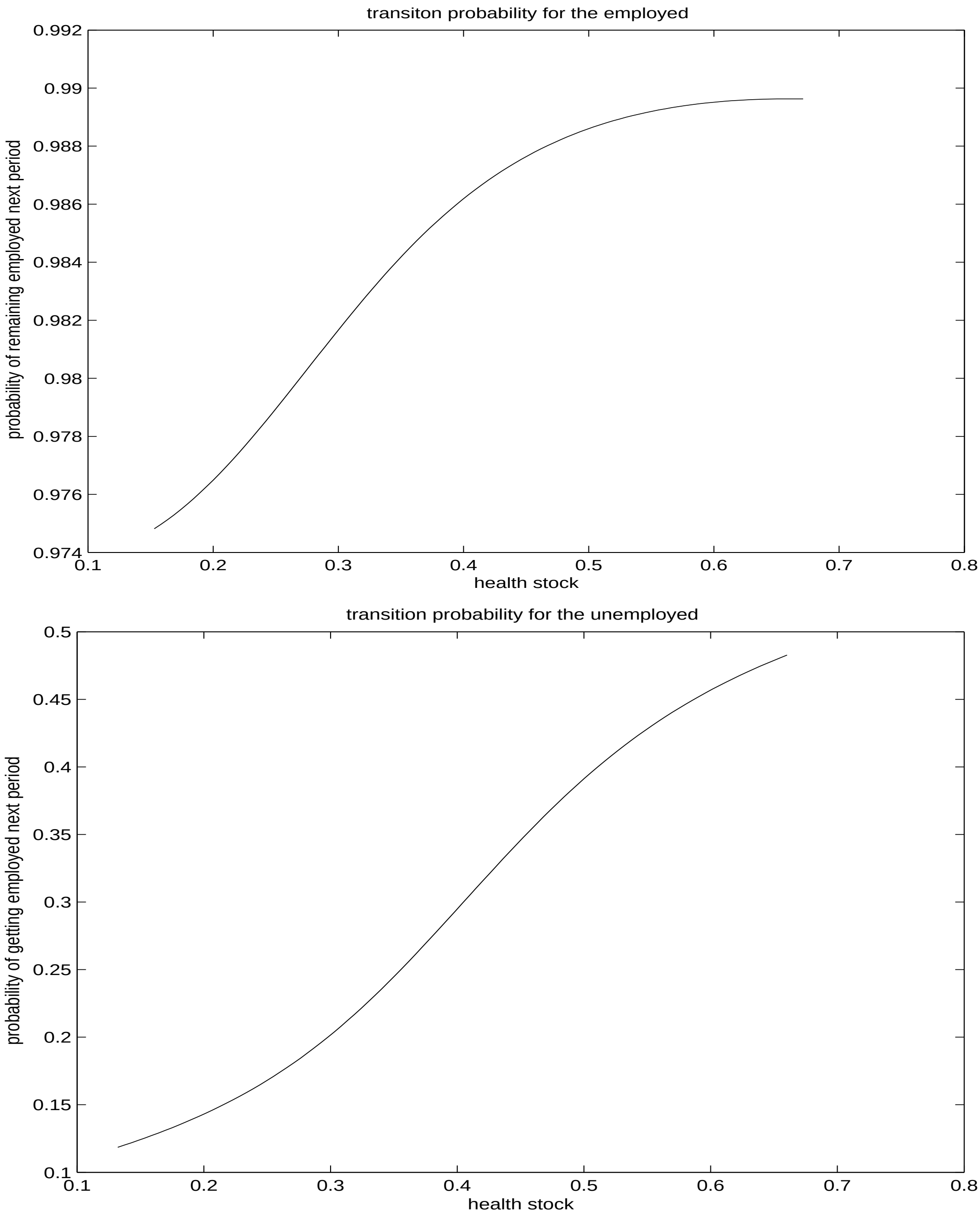

Figure 3.1: The employment transition probabilities for employed and unemployed individuals

we can see that everything else being equal, a healthier agent has a higher chance of getting the job and the agent who had a job last period has a better chance of being employed this period.

In MEPS, employment status is surveyed every half a year but medical expenditure is the expenditure accumulated within one year. In order to create yearly data, the individual is assumed to be employed as long as he is employed in either half of the year. This will result in an underestimate of the impact of health status on individual's employment status and increase the persistence of employment status in the top graph of Figure 3.1. It would be better if we have monthly or quarterly data in order to estimate these transition functions more accurately. However, it won't have a big impact on the robustness of the structural parameter estimates later on since the same sort of bias applies to the reduced form data regression.

3.4.3 Health Insurance Choice

Compensating wage theory predicts that workers receiving more generous fringe benefits are paid a lower wage than comparable workers who prefer fewer fringe benefits. In this paper, there are two compensation packages available to the employed individual(with health insurance or without). If the worker chooses the package with health insurance, his annual income is y_c, and he pays the health insurance premium of p_e. As a result, his disposable income is $y_c(1-\tau-p_e)$, where τ is the income tax he has to pay. This paper assumes that employee portion of health insurance premium is not tax deductible in the benchmark model. If the worker chooses the package without health insurance, his annual income is $y_c + R$. As a result, his disposable income is $(y_c + R)(1 - \tau)$. R is the wage differential between the packages with and without health insurance.

If the agent is unemployed, his disposable income is just θy, where θ is the replacement ratio of unemployment insurance and y is the mean wage of the employed individuals. This paper assumes that unemployed individuals all have health insurance provided by some government program for simplicity. This assumption shouldn't have a big impact on the result of this analysis for the following reason. A big portion (about 64.6 percent, see Table 1) of unemployed individuals has some form of public health insurance. Only 12.44 percent of the unemployed individuals has no health insurance.

Table 3.1: Working Aged Individuals with Selected Sources of Health Insurance from MEPS Data:

	Employed		**Unemployed**		**Total**	
	obs.	**per**	**obs.**	**per**	**obs.**	**per**
Total	7097	100	418	100	7515	100
Employment	5377	75.76	82	19.62	5459	72.64
Public	339	4.78	270	64.6	609	8.1
Private	278	3.92	14	3.35	292	3.89
Uninsured	1103	15.54	52	12.44	1155	15.37

Table 3.1 presents the sources of health insurance for the employed and unemployed working aged individuals from MEPS. *Total* gives us how

many observations are in each sample (employed, unemployed and in total). *Employment*, *Public*, *Private* and *Uninsured* give us how many observations have employer-provided health insurance, public insurance, private insurance (other than employer-provided) and don't have health insurance at all respectively. *Employed*, *Unemployed* and *Total* in the first row summarize the number of observations falling into each category described above for employed individuals, unemployed individuals and in total respectively. There are 7515 observations in the sample. 7097 are employed and 418 are unemployed. For the employed, 75.76 percent has employer-provided health insurance and 15.54 percent is uninsured. The percentages of employed individuals who have public and private health insurance are 4.78 and 3.92 respectively. For the unemployed, about 64.6 percent has some form of public health insurance. Only 12.44 percent of the unemployed individuals has no health insurance. The percentages of unemployed individuals who have employment-based (COBRA) health insurance and private health insurance are 19.62 and 3.35 respectively.

3.4.4 Timing and Defining Maximization Problem

The timing of the model is the following: The agent enters period t with a health stock of h_t and then his employment status is revealed. If he is employed, he then decides whether to buy health insurance or not. Then a health shock ϕ_t is realized. After observing his health shock, the agent chooses m_t either to recover from a bad shock and/or improve his health. Then he enters period t+1 with a health stock of h_{t+1} which evolves according to equation (3.3).

The maximization problem can be written as a dynamic programming problem. Note that the state variables are health stock h and employment status s.

The dynamic programming problem for the employed individuals is:

$$V(e,h) = max\{V_c(e,h), V_{nc}(e,h)\} \tag{3.5}$$

subject to equation (3.2), where

$$V_c(e,h) = \sum_{\phi} \mu(\phi)\{\max_m u[y_c(1-\tau-p_e) - \alpha m, h] + \beta \sum_{s'} \pi(s'|e,h')V(s',h')\}$$

is the value of choosing health insurance coverage but getting a lower wage, and

$$V_{nc}(e,h) = \sum_{\phi} \mu(\phi)\{\max_m u[(y_c+R)(1-\tau) - m, h] + \beta \sum_{s'} \pi(s'|e,h')V(s',h')\}$$

is the value of choosing without health insurance coverage but getting a higher wage.

The dynamic programming problem for the unemployed individuals is:

$$V(u,h) = \sum_{\phi} \mu(\phi)\{\max_m u[\theta y - \alpha m, h] + \beta \sum_{s'} \pi(s'|u,h')V(s',h')\} \tag{3.6}$$

subject to

$$m \geq 0 \tag{3.7}$$

3.5 Parameterizations

This section first describes how those parameter values in the theoretical model are chosen and what parameters need to be estimated. Then it presents the steps of this indirect inference estimation strategy. A descriptive model which is the basis for this indirect inference procedure is then formulated at the end of this section.

3.5.1 Parameters in the Structural Model

The utility function used in the computation has the following form:

$$U(c,h) = \frac{(c^{1-\sigma}h^{\sigma})^{1-\rho} - 1}{1-\rho} \tag{3.8}$$

where $0 \leq \sigma < 1$ and $\rho > 0$ guarantee that $u_c > 0$, $u_h > 0$, $u_{cc} < 0$ and $u_{hh} < 0$ but the sign of u_{ch} depends on whether ρ is greater or less than 1.

The time period in the model is one year. The health stock is set between 0 and 1 to match the data. β is set to 0.95. δ is the health depreciation rate in the health evolution equation. δ is set to 0.01 based on the following calculation. The lowest health level in the data is 11.73 and the highest is 67.24. This paper assumes that human beings can live up to 150 years if without any accident or illness. Then, $67.24(1-\delta)^{150} = 11.73$. This gives us the value of δ. If this paper assumes that human beings can live up to 100 years without any accident or illness, δ is calculated to be 0.017. This does not affect the main results of this paper. The coinsurance rate α is set to be 0.19 which is the average coinsurance rate from MEPS.

The wage rate with health insurance coverage y_c is normalize to 1. The wage premium without health insurance coverage R is set to 0.1 and the employee portion of health insurance premium p_e is set to 0.02. The wage differential between the compensation package without insurance coverage and with insurance coverage is 12 percent in this paper (p_e+R). In the literature of compensating wage differentials for health insurance benefit provision, economic theory suggests that what matters to profit maximization firms is the value of the total compensation package that they must offer to attract labor services. To remain competitive, the firm must reduce wages by 1 dollar for each 1 dollar increase in health insurance expenditures. However, the empirical estimation often results in a positive relationship between health expenditures and wages which contradicts the theory. This is mainly because of the lack of the suitable data (Smith and Ehrenberg, 1983) since we only observe individuals who end up having health insurance or not but we don't observe the alternative forgone offers. With the usual data, we are comparing the insurance choices of different individuals. However, the workers with higher abilities will work in the firm with both higher wage and better health insurance plan. MaCurdy and Rapoport (2003) is able to identify the health insurance and wage tradeoff for the low-skilled workers in the US which is about 9 percent. The sample of that study consists of the individuals who switch between jobs with and without health insurance. Olson(2002) finds that wives with own employer health insurance accept a wage about 20 percent lower than what they would have received working in a job without benefits. That study tests wage differential theory for employer-provided health insurance by modeling the wages of married women employed full-time in the labor market. Husband's union status, husband's firm size, and husband's health coverage through his job are used as instruments for his wife's own employer health benefits. The wage differential in this paper is set between the above two estimates. Later version of this paper might estimate this wage differential as well.

This paper takes policy parameters θ and τ as exogenous. θ is set to 0.25 for the following reason. The U.I. replacement ratio in the United States is about 0.5 which usually lasts for six months. Since the unemployment duration in this model is assumed to be 1 year so the U.I. replacement ratio is set to be 0.25 in this paper and τ is set to 0.15 which is the average federal tax rate in the United States.

The parameters need to be estimated by Indirect Inference are $\Theta = \{\sigma\ \rho\ a\ b\ z\}$. σ and ρ are preference parameters. a and b are parameters in the health stock evolution equation. z is the health shock parameter.

3.5.2 Indirect Inference

The structural model parameters are estimated using the method of indirect inference. For arbitrary values of the vector of parameters Θ, the dynamic programming problem is solved and policy functions are generated. Using these policy functions, the decision rule is simulated to create a simulated version of the data to match. One then chooses a descriptive statistical model that provides a rich description of the patterns of covariation in the data. Such a descriptive model can be estimated on both the simulated data from the structural model, and on the actual observed data. This then gives us two sets of coefficients to match, $\Psi^s(\Theta)$ and Ψ^d.

The estimate $\hat{\Theta}$ is pinned down by minimizing the weighted distance between the actual and simulated coefficients from the descriptive models. Formally, it solves

$$\pounds(\Theta) = \min_{\Theta}[\Psi^d - \Psi^s(\Theta)]' W [\Psi^d - \Psi^s(\Theta)] \tag{3.9}$$

where W is a weighting matrix. The method of indirect inference will generate a consistent estimate of θ. The weighting matrix,W, is constructed as the inverse of the variance-covariance matrix of the coefficients estimated from the MEPS.

Since the $\Psi^s(\Theta)$ function is not analytically tractable, the minimization is performed using numerical techniques. A simulated annealing algorithm is used to perform the optimization in order to obtain the global minimum in parameter space independent of starting values.

3.5.3 Descriptive Model

The descriptive model consists of two linear equations which are extensively estimated in the literature of health economics.[3] These equations are

[3] Currie and Madrian (1999) has a detailed review of them. Stratmann (1999) estimates the effect of doctor visits on work day loss using the types of health insurance as instruments for doctor visits.

health equation and medical expenditure equation. They take the following forms:

$$\begin{aligned} h_{t+1} &= \alpha_1 m_t^o + \alpha_2 h_t + \alpha_3 h_t m_t^o + X_1\alpha_4 + \varepsilon_{1t} \\ m_t^o &= \gamma_1 h_t + \gamma_2 s_t + X_2\gamma_3 + \varepsilon_{2t} \end{aligned} \tag{3.10}$$

where h is individual's health, s is individual's employment status and m is individual's out-of-pocket medical expenditure.[4] X_1 and X_2 are control variables in these equations. Since these control variables are not modeled in the structural model, it is assumed that the agents in the structural model are homogenous in them. The MEPS coefficients which are going be matched by the simulated coefficients are $\{\alpha_1\ \alpha_2\ \alpha_3\ \gamma_1\ \gamma_2\}$. Tables 3.2 and 3.3 provide the definitions and the summary statistics of all the variables in these equations.

In the health equation, medical expenditure is expected to have a positive effect on health but the coefficient in the OLS regression is usually negative. Last period health is expected to have a positive effect on this period health.

In the medical expenditure equation, conditional on the same employed status, healthier individuals are expected to spend less on their medical bill and therefore, the coefficients is expected to be negative. Conditional on the same health level, the employed individuals are expected to spend more on

[4]Here, out-of-pocket medical expenditure is in the form of the percentage of income in order to match the simulated data from the structural model. In order to use the medical expenditure share, the individuals without any income or with the medical expenditure share greater than 1 are out of sample.

their medical care and therefore the coefficient is expected to be positive but in the reduced form regressions, the coefficient is significantly negative.

The coefficients on the control variables in the two equations all have their expected signs. These two equations are estimated separately by OLS regression. The results are reported in Table 3.4.

There are six parameters to pin down in the structural model but there are only five coefficients from the descriptive model. The additional moment used is the health insurance take-up rate for the employed individuals from MEPS (Table 3.1) which is 0.8446. As a result,

$$\Psi^d = \{-0.4249\ 0.5325\ 0.5508\ -0.0918\ -0.0437\ 0.8446\}.$$

3.6 Estimation Results

The estimation procedure described in the previous section gives us the following results of the structural parameters and the coefficients from the simulated data:

From Table 3.6, we can see that the simulated coefficients match the MEPS coefficients quite well. Interestingly, the simulated α_1, the coefficient of medical expenditure in the health equation, is negative even though in the health evolution equation of the structural model, medical expenditure only

Table 3.2: Variable definitions:

Dependent variables	
$health_{t+1}$	physical component summary index* at the end of period t, 0 is the lowest heath level, 100 is the highest
$medical_t$	total amount of out-of-pocket medical expenditure
Independent Variables appearing in all the equations	
age	Individual's actual age
sex	1 if male, 0 otherwise
race	1 if white, 0 otherwise
mar	1 if married, 0 otherwise
Additional independent variables appearing in the health equation	
$health_t$	physical component summary index at the end of period t-1, 0 is the lowest heath level, 100 is the highest
Additional independent variables appearing in the medical expenditure equation	
$employed_t$	employment status at period t: 1 if in labor force, 0 otherwise

*Refer Appendix 1 for the description of health measure used in this paper.

Table 3.3: Descriptive statistics of the sample:

Variables	Sample Mean	Standard dev.	Minimum	Maximum
Dependent variables				
$health_{t+1}/100$	0.51	0.08	0.1173	0.6724
$medical_t^o/income_t$	0.02	0.06	0	0.9983
Independent Variables appearing in all the equations				
$age/100$	0.41	0.11	0.19	0.64
sex	0.49	0.50	0	1
$race$	0.81	0.39	0	1
mar	0.62	0.49	0	1
Additional independent variables appearing in the health equation				
$health_t/100$	0.51	0.09	0.1321	0.6713
Additional independent variables appearing in the medical expenditure equation				
$employed_t$	0.94	0.23	0	1

Table 3.4: Estimation results of OLS estimation

Variables	Coefficients of The Variables	Std
Health equation		
$cons$	0.2646	0.0066
$medical^o_t/income_t$	-.4249	0.053
$health_t/100$	0.5325	0.0103
$(health_t/100)(medical^o_t/income_t)$	0.5508	0.1139
$age/100$	-.0735	0.0070
sex	0.0047	0.0016
$race$	0.0028*	0.0020
mar	0.0042	0.0016
R^2	0.3962	
Medical expenditure equation		
$cons$	0.0950	0.0059
$health_t/100$	-.0918	0.0087
$employed_t$	-.0437	0.0032
$age/100$	0.0509	0.0064
sex	-.0138	0.0014
$race$	0.0057	0.0018
mar	-.0050	0.0015
R^2	0.0863	
$No\ of\ observations$	7515	

*the coefficient is not significant at 5 percent significant level.

Table 3.5: Structural Parameter Estimates:

	σ	ρ	a	b	z
Estimated Value	0.1839	1.7239	0.1398	0.8828	0.6679
Standard Error	0.0021	0.0271	0.0009	0.0014	0.0009

Table 3.6: Coefficients and Moments from the MEPS and Simulated Data:

	α_1	α_2	α_3	γ_1	γ_2	*turate*	$\pounds(\Theta)$
MEPS	-0.4249	0.5325	0.5508	-0.0918	-0.0437	0.8446	
Simulated	-0.1508	0.4768	-0.0247	-0.2857	-0.0724	0.8993	566

impacts the health in the positive way because parameters a and b are both positive. This is caused by the endogeneity problem in the reduced form health production function regression. Something in the error term might be correlated with the medical expenditure in the reduced form health production function (Refer Han (2005)). The fact reflects that reduced form regression results sometimes give us the wrong information without knowing the structural model behind it because of some correlation issues. The only simulated coefficient which failed to match the data coefficient is the coefficient on cross term of medical expenditure and heath status. A partial explanation for the lack of fit of this moment is the low weight placed on the relative standard deviation compared to weights on the other moments, as determined by the inverse of the variances of the respective moments from the MEPS.

From a statistical perspective, the model is rejected since the reported values of $\pounds(\Theta)$ are still high compare to the cut off value. However, in this setting, this reflects the fact that the coefficients are calculated from a very large panel data set, implying very small standard deviations of the coefficients (and a very large W). Given how precisely the micro coefficients are estimated from the actual data, virtually any model would be formally rejected with even very modest deviations of the simulated coefficients from the actual coefficients. As we have emphasized above, the fit of the model in the last line of Table 3.6 is actually quite good in terms of matching the data coefficients on both a qualitative and quantitative basis.

3.7 Welfare Analysis

3.7.1 Policy Experiment and Welfare Measure

With the structural parameters being estimated, this paper is able to simulate individuals' policy function and decision rules after imposing the mandatory employer-provided health insurance. For example, the employed individuals don't have the no-insurance option anymore. The wage now is $y_c - p_e$. Then, the welfare cost is measured by μ where $\bar{V}^b(y^d(1-\mu)) = \bar{V}^a(y^d)$. $\bar{V}^b$ is the average utility before the mandate and $\bar{V}^a$ is the average utility after the mandate. If μ is negative, then it is welfare gain instead of cost. If μ is positive, it reflects the percentage of the income all the agents in the economy are willing to give up in order to avoid this mandate. If μ is negative, it reflects the percentage of the income all the agents in the economy are willing to pay to for this mandate.

3.7.2 Measuring the Welfare Cost

I assume that immediately after this mandate, the firm cancels the no-insurance option and adjust the wage to $y_c - p_e$ and the health insurance is in the total compensation package to the worker. In the short run, no budget balance rule is imposed and therefore,we don't worry about the tax issue for now. As a result, this paper can only measure the cost of this mandate for now. There are two regions we need to consider: not affected region and worse-off region. The not affected region includes those individuals who are employed and already chose the health insurance option before the mandate. Their disposable income remains the same as before and therefore, their decision rules on the medical expenditure choices. The worse-off regions includes those individuals who are employed but didn't choose the health insurance before the mandate. It is not hard to see why this group is worse off since they are forced to choose something which they did't want before the mandate. The μ calculated from the simulation is 0.007. In another word, the agents in this economy are willing to give up 0.7 percent of GDP in total to avoid this mandate in the short run.

In the long run, even though this mandate eliminates individuals' choices, it does guarantee a healthier population and decreases the society's burden. As a result, the welfare change could go either way. Less people will collect unemployed insurance and more people are paying the income tax. This could result in a decrease in the tax rate or an increase in the unemployment benefits. This paper is only dealing with the partial equilibrium for now and measuring the cost of mandatory health insurance. Later version of this paper might consider simulating the general equilibrium effect of this mandate.

3.8 Conclusion

This paper measures the cost of mandatory employer-provided health insurance policy by comparing the nation's welfare with and without this policy. This paper allows the individuals to choose their own medical expenses to improve their health status and in turn increase their job opportunities when facing the unexpected health shocks in a dynamic structural model. In this model, they also choose to have the health insurance coverage or not according to their own characteristics. The structural parameters in the model are estimated by matching the model's implications with individual observations from the Medical Expenditure Panel Survey (MEPS) as part of a minimum distance estimation routine. The mandate further distort the market equilibrium from the competitive equilibrium. It suppresses individuals' choices and will result in inefficiency. The welfare cost measured from this paper is 0.7 percent of GDP.

Later version of the paper will simulate the general equilibrium effect of this mandate to incorporate the welfare gain from this mandate.

Appendix

Appendix 1

Appendix for Chapter 1

Estimation results of OLS estimation:

Variables in The Equations	Coefficients of The Variables	Standard Error
Health equation		
$cons$	2.3421	0.0604
$employed_t$	0.1923	0.0179
$medical_t/1000$	-.2047	0.0216
$age/10$	-.0829	0.0065
sex	0.0159*	0.0150
$race$	0.0363	0.0187
mar	0.0521	0.0155
edu	0.1116	0.0177
$smoke$	-.0647	0.0175
$phyact$	0.0987	0.0147
$seatbelt$	0.0293*	0.0174
$health_{t-1}/10$	0.5493	0.0093
$health_{t-1}/10medical_t/1000$	0.0271	0.0046
R^2	0.4814	
Labor force participation equation		
$cons$	-.2002	0.0394
$health_t$	0.0359	0.0031
age	0.1346	0.0179
sex	0.0052*	0.0072
$race$	0.0063*	0.0070
mar	0.0309	0.0085
edu	0.0212	0.0067
$age2$	-.0179	0.0021
$numchi$	-.0050*	0.0054
$child5$	-.0402	0.0177
$employed_{t-1}$	0.7357	0.0068
$spousein$	-.0006	0.0001
$sexnchi$	0.0157	0.0052
$sexchi5$	0.0207	0.0158
$marnchi$	-.0131	0.0061
$marchi5$	0.0080*	0.0193
R^2	0.6502	

Estimation results of OLS estimation(Continued):

Variables in The Equations	Coefficients of The Variables	Standard Error
Medical expenditure equation		
cons	1.1322	0.0885
$employed_t$	-.0721	0.0290
$health_t$	-.2405	0.0133
age	0.1360	0.0102
sex	-.1622	0.0236
race	0.1901	0.0296
mar	-.0770	0.0287
edu	0.2050	0.0285
hmo	-.1492	0.0260
gatekeeper	-.0478*	0.0442
ppo	0.0485*	0.0334
spousein	0.0016	0.0005
R^2	0.1044	
No of observations	8896	

Bibliography

[1] Abdulkadiroglu, A., B. Kuruscu and A. Sahin, "Unemployment Insurance and the Role of Self-Insurance," *Review of Economic Dynamics*, Vol. 5, July 2002, 681-703.

[2] Adda, J. and R. Cooper, *Dynamic Economics: Quantitative Methods and Applications*, The MIT Press, Cambridge, Massachusetts, 2003.

[3] Amemiya, T., "The Estimation of a Simultaneous Equation Generalised Probit Model." *Econometrica* 46: 1193-1205, 1978.

[4] Avery, J, E. D. Grande and A. Taylor, *Quality of Life in South Australia as Measured by the SF12 Health Status Questionnaire.* Population Research and Outcome Unit, Department of Human Services, South Australia, 2004.

[5] Blau, D. M. and D. B. Gilleskie, "Health Insurance and Retirement of Married Couples," University of North Carolina at Chapel Hill working paper, 2005.

[6] Cai, L. and G. Kalb, "Health Status and Labour Force Participation: Evidence from Australia." Health Economics, 15:241-261, 2006.

[7] Chatterjee, S., D. Corbae, M. Nakajima and J. Rios-Rull, "A Quantitative Theory of Unsecured Consumer Credit with Risk of Default," Federal Reserve Bank of Philadelphia Working Paper No. 05-18.

[8] Cropper, M. L., "Health, Investment in Health, and Occupational Choice," *The Journal of Political Economy*, Vol.85, No. 6 (Dec., 1977), 1273-1294.

[9] Currie, J. and Madrian, B. C., Health, Health Insurance and the Labor Market. *Handbook of Labor Economics*. Ashenfelter, O. and Card, D., Elsevier Science B.V. 3: 3310-3415, 1999.

[10] De Nardi, M., E. French and J.B. Jones, "Differential Mortality, Uncertain Medical Expenses, and the Saving of Elderly Singles," Federal Reserve Bank of Chicago Working Paper No. 05-13.

[11] Fronstin, Paul."Sources of Health Insurance Coverage and Characteristics of the Uninsured: Analysis of the March 2003 Current Population Survey." *BERI Issue Brief* no. 276 (Employee Benefit Research Institute, December 2004).

[12] Greene, W. H., *Econometric Analysis* (Fifth Edition). New York: Macmillan Publishing Company, 1993.

[13] Gerdtham, U. and M. Johannesson, "A Note on the Effect of Unemployment on Mortality." *Journal of Health Economics*, Vol. 22, 2003, 505-518.

[14] Gourieroux, C. and A. Monfort, *Simulation-Based Econometric Methods*, Oxford University Press Inc., New York, 2002.

[15] Grossman, M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, Vol.80, No.2, 1972, 223-255.

[16] Grossman, M. and T. J. Joyce, "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City," *Journal of Political Economy*, Vol. 98, No.5, Part 1, Oct. 1990, 983-1007.

[17] Grossman, M., *The Demand for Health: A Theoretical and Empirical Investigation*, Columbia University Press for the National Bureau of Economics Research, New York, 1972.

[18] Grossman, M.. The Human Capital Model of The Demand for Health (1999), *NBER working parper* No. 7078.

[19] Gruber, J. and A. Krueger,"The Incidence of Mandated Employer-Provided Insurance: Lessons from Workers' Compensation Insurance," NBER working paper no. 3557.

[20] Hall, R. E., and C.I. Jones, "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, forthcoming.

[21] Han, X.,"A Medical Expenditure Puzzle," working paper of University of Texas at Austin 2005.

[22] Hansen, G. D., and A. Imrohoroglu, "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard," *Journal of Political Economy*, Vol.100, 1992, 118-142.

[23] Hubbard, G. R., J. Skinner and S. P. Zeldes, "Expanding the Life-Cycle Model: Precautionary Saving and Public Policy," *American Economic Review*, Vol.84, 1994, 174-179.

[24] Hubbard, G. R., J. Skinner and S. P. Zeldes, "Precautionary Saving and Social Insurance," *Journal of Political Economy*, Vol.103(2), 1995, 360-399.

[25] Hurst, N. P., D. A. Ruta and P. Kind, "Comparison of the Mos Short Form-12(SF12) Status Questionnaire with the SF36 in Patients with Rheumotoid Arthritis." *British Journal of Rheumatology* 37: 862-869, 1998.

[26] Keane, M. and A. A. Smith, Jr., "Generalized Indirect Inference for Discrete Choice Models." *manuscript* (Yale University).

[27] Kotlikoff, L. J., "Health Expenditures and Precautionary Savings," In *What Determines Savings?* Massachusetts, MIT Press,1989.

[28] Kreider, B., "Latent Work Disability and Reporting Bias. " *The Journal of Human Resources* 34(4): 734-769, 1999.

[29] Kydland, Finn E., and Prescott, Edward C., "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (1982), 1345-1370.

[30] Lee, L., "Health and Wages: A Simultaneous Equation Model with Multiple Discrete Indicators." *International Economic Review* 23: 199-221, 1982.

[31] Luo, X., L. George, Mandy, Kakouras, Ikey, C. L. Edwards, Pietrobon, Ricardo, Richardson, William, Hey and Lloyd, "Reliability, Validity, and Responsiveness of the Short Form 12-Item Survey (SF-12) in Patients with Back Pain." *Spine* 28(15):1739-1745, 2003.

[32] MaCurdy, T. and D. Rapoport, "Understanding Health-Insurance Coverage of Low-Skilled Workers in the US and in California: Trading Wages for Insurance." *SPHERE Institure*, Burlingame, CA, January 2003.

[33] Maddala, G., *Limited Dependent and Qualitative Variables in Econometrics.* New York: Cambridge University Press, 1983.

[34] Melville, M. R., M. A. Lari, N. Brown, T. Young and D. Gray, "Quality Of Life Assessment Using The Short Form 12 Questionnaire Is As Reliable And Sensitive As The Short Form 36 In Distinguishing Symptom Severity In Myocardial Infarction Survivors." *Heart* 89:1445-1446, 2003.

[35] Newhouse, J. P., "Determinants of Days Lost from Work due to Sickness." *Empirical Studies in Health Economics*, edited by H.E. Klarman. Baltimore, 1970.

[36] Olson, C., "Do Workers Accept Lower Wages in Exchange for Health Benefits?" *Journal of Labor Economics*, Vol.20, No. 2, Part 2, (April 2002), 91-114.

[37] Palumbo, M. G., "Uncertain Medical Expenses and Precautionary Saving

Near the End of the Life Cycle," *Review of Economic Studies*, Vol.66, 395-421, 1999.

[38] Rosenzweig, M. R. and K. L. Wolpin, "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, Vol. 38, No. 4, 827-874, Dec., 2000.

[39] Rosenzweig, M. R. and T. P. Schultz, "Estimating a Household Production: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy*, Vol. 91, No.5, Oct. 1983, 723-746.

[40] Rust, J. and C. Phelan, "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, Vol. 65, 781-831,1997.

[41] Smith, R. S. and R. G. Ehrenberg, "Estimating Wage-fringe Trade-offs: Some Data Problems," *The Measurement of Labor Cost*, University of Chicago Press, Chicago, IL, pp.347-367, 1983.

[42] Stern, S., "Measuring the Effect of Disability on Labour Force Participation." *The Journal of Human Resources* 24(3): 361-395, 1989.

[43] Stewart, J. M., "The Impact of Health Status on the Duration of Unemployment Spells and the Implications for Studies of the Impact of Unemployment on Health Status." *Journal of Health Economics*, Vol. 20, 2001, 781-796.

[44] Stratmann, T., "What Do Medical Services Buy? Effects of Doctor Visits on Work Day Loss." *Eastern Economic Journal* 25(1), 1999.

[45] Summers, L., "Some Simple Economics of Mandated Benefits," *American Economic Association Papers and Proceedings*, 79, 1989, 177-183.

[46] Wagstaff, A., "The Demand for Health: An Empirical Reformulation of the Grossman Model", *Health Economics*, Vol. 2, 1993, 189-198.

[47] Wilson, S. H. and Walker, G. M., "Unemployment and Health: A Review." *Public Health* 107: 153-162, 1993.

[48] Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, 2001.

Vita

Xiaoshu Han was born in YuShu, JiLin province, P.R. China on 14 March 1977. She received the Bachelor of Arts degree in English with International Business minor from Beijing Foreign Studies University in 1999 and was employed as a strategy analyst in Tingyi (Cayman Island) Holding Corp in Tianjin, P.R. China for one year. She separated from Tingyi (Cayman Island) Holding Corp in 2000 and entered University of New York at Binghamton to pursue her graduate study in economics in August, 2000 and received her Master of Arts degree in May 2002. She then applied to the University of Texas at Austin for enrollment in their doctoral program in Economics to further pursuing her dream to be a college teacher in the future. She was accepted and started doctoral studies in August, 2002.

Permanent address: 1510 W. N. Loop Blvd. Unit 114
Austin, Texas 78756

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.